Amendments to the Claims

Please amend the claims as follows:

1-25. (Cancelled).

26. (Previously Presented): A method for building a decision tree from an input data set, the input data set comprising records and associated attributes, the attributes including a class label attribute for indicating whether a given record is a member of a target class or a non-target class, the input data set being biased in favor of the records of the non-target class, the decision tree comprising a plurality of nodes that include a root node and leaf nodes, said method comprising the steps of:

constructing the decision tree from the input data set, including the step of partitioning each of the plurality of nodes of the decision tree, beginning with the root node, based upon multivariate subspace splitting criteria;

computing distance functions for each of the leaf nodes;

identifying, with respect to the distance functions, a nearest neighbor set of nodes for each of the leaf nodes based upon a respective closeness of the nearest neighbor set of nodes to a target record of the target class; and

classifying and scoring the records, based upon the decision tree and the nearest neighbor set of nodes;

wherein said partitioning step comprises the step of:

for a current leaf node from among the leaf nodes of the decision tree,

computing a lowest value of a gini index achieved by univariate-based partitions on each of a plurality of attribute lists included in the current leaf node; and

wherein the gini index is equal to 1 - $(P\_n)^2$ - $(P\_p)^2$, P_n being a percentage of the records of the non-target class in the input data set and P_p being a percentage of the records of the target class in the input data set.

27. (Previously Presented): A method for building a decision tree from an input data set, the input data set comprising records and associated attributes, the attributes including a class label attribute for indicating whether a given record is a member of a

target class or a non-target class, the input data set being biased in favor of the records of the non-target class, the decision tree comprising a plurality of nodes that include a root node and leaf nodes, said method comprising the steps of:

constructing the decision tree from the input data set, including the step of partitioning each of the plurality of nodes of the decision tree, beginning with the root node, based upon multivariate subspace splitting criteria;

computing distance functions for each of the leaf nodes;

identifying, with respect to the distance functions, a nearest neighbor set of nodes for each of the leaf nodes based upon a respective closeness of the nearest neighbor set of nodes to a target record of the target class; and

classifying and scoring the records, based upon the decision tree and the nearest neighbor set of nodes;

wherein said partitioning step comprises the step of:

for a current leaf node from among the leaf nodes of the decision tree,

computing a lowest value of a gini index achieved by univariate-based partitions on each of a plurality of attribute lists included in the current leaf node; and

wherein the percentage of the records P_p in the input data set is equal to W_p * n_p / (W_p * n_p + n_n), W_p being a weight of the records of the target class in the input data set, n_p and n_n being a number of the records of the target class and a number of the records of the non-target class in the current leaf node, respectively.

28. (Previously Presented): A method for building a decision tree from an input data set, the input data set comprising records and associated attributes, the attributes including a class label attribute for indicating whether a given record is a member of a target class or a non-target class, the input data set being biased in favor of the records of the non-target class, the decision tree comprising a plurality of nodes that include a root node and leaf nodes, said method comprising the steps of:

constructing the decision tree from the input data set, including the step of partitioning each of the plurality of nodes of the decision tree, beginning with the root node, based upon multivariate subspace splitting criteria;

computing distance functions for each of the leaf nodes;

identifying, with respect to the distance functions, a nearest neighbor set of nodes for each of the leaf nodes based upon a respective closeness of the nearest neighbor set of nodes to a target record of the target class; and

classifying and scoring the records, based upon the decision tree and the nearest neighbor set of nodes;

wherein said partitioning step comprises the step of:

for a current leaf node from among the leaf nodes of the decision tree,

computing a lowest value of a gini index achieved by univariate-based partitions on each of a plurality of attribute lists included in the current leaf node; and

wherein said partitioning step further comprises the steps of:

detecting subspace clusters of the records of the target class associated with the current leaf node;

computing the lowest value of the gini index achieved by distance-based partitions on each of the plurality of attribute lists included in the current leaf node, the distance-based partitions being based on distances to the detected subspace clusters;

partitioning pre-sorted attribute lists included in the current node into two sets of ordered attribute lists based upon a greater one of the lowest value of the gini index achieved by univariate partitions and the lowest value of the gini index achieved by distance-based partitions; and

creating new child nodes for each of the two sets of ordered attribute lists; and

wherein said detecting step comprises the steps of:

computing a minimum support (minsup) of each of the subspace clusters that have a potential of providing a lower gini index than that provided by the univariate-based partitions;

identifying one-dimensional clusters of the records of the target class associated with the current leaf node;

beginning with the one-dimensional clusters, combining centroids of K-dimensional clusters to form candidate (K+1)-dimensional clusters;

identifying a number of the records of the target class that fall into each of the (K+1)-dimensional clusters;

pruning any of the (K+1)-dimensional clusters that have a support lower than the minsup.

29. (Previously Presented): The method of claim 28, wherein the support of a subspace cluster is denoted as n_p'/n_p, n_p' being a number of the records of the target class in the subspace cluster, and n_p being a total number of the records of the target class in the current leaf node.

30. (Previously Presented): The method of claim 29, wherein the minsup is denoted as (2q-2q^2-G_best)/(2q-2q^2-qG_best), G_best being a smallest gini index given by the univariate-based partitions, q being n_p/n_n, and n_n being a total number of the records in the current leaf node.

31. (Previously Presented): The method of claim 28, wherein said step of identifying the one-dimensional clusters of the records of the target class comprises the steps of:

dividing a domain of each dimension of a data set associated with the current leaf node into a predetermined number of equal-length bins;

identifying all of the records of the target class falling into each of the predetermined number of equal-length bins; and

for each of a current dimension of the data set associated with the current leaf node,

constructing a histogram for the current dimension; and

identifying clusters of records of the target class on the current dimension, using the histogram.

32. (Previously Presented): A method for building a decision tree from an input data set, the input data set comprising records and associated attributes, the attributes including a class label attribute for indicating whether a given record is a member of a target class or a non-target class, the input data set being biased in favor of the records of the non-target class, the decision tree comprising a plurality of nodes that include a root node and leaf nodes, said method comprising the steps of:

constructing the decision tree from the input data set, including the step of partitioning each of the plurality of nodes of the decision tree, beginning with the root node, based upon multivariate subspace splitting criteria;

computing distance functions for each of the leaf nodes;

identifying, with respect to the distance functions, a nearest neighbor set of nodes for each of the leaf nodes based upon a respective closeness of the nearest neighbor set of nodes to a target record of the target class; and

classifying and scoring the records, based upon the decision tree and the nearest neighbor set of nodes;

wherein said partitioning step comprises the step of:

for a current leaf node from among the leaf nodes of the decision tree,

computing a lowest value of a gini index achieved by univariate-based partitions on each of a plurality of attribute lists included in the current leaf node; and

wherein said partitioning step further comprises the steps of:

detecting subspace clusters of the records of the target class associated with the current leaf node;

computing the lowest value of the gini index achieved by distance-based partitions on each of the plurality of attribute lists included in the current leaf node, the distance-based partitions being based on distances to the detected subspace clusters;

partitioning pre-sorted attribute lists included in the current node into two sets of ordered attribute lists based upon a greater one of the lowest value of the gini index achieved by univariate partitions and the lowest value of the gini index achieved by distance-based partitions; and

creating new child nodes for each of the two sets of ordered attribute lists; and

wherein said step of computing the lowest value of the gini index achieved by distance-based partitions comprises the steps of:

identifying eligible subspace clusters from among the subspace clusters, an eligible subspace cluster having a set of clustered dimensions such that only less than all of the clustered dimensions in the set are capable of being included in another set of clustered dimensions of another subspace cluster;

selecting top-K clusters from among the eligible subspace clusters, the top-K clusters being ordered by a number of records therein;

for each of a current top-K cluster,

computing a centroid of the current top-K cluster and a weight on each dimension of the current top-K cluster; and

computing the gini index of the current top-K cluster, based on a weighted Euclidean distance to the centroid; and

recording a lowest gini index achieved by said step of computing the gini index of the current top-K cluster.

33. (Previously Presented): A method for building a decision tree from an input data set, the input data set comprising records and associated attributes, the attributes including a class label attribute for indicating whether a given record is a member of a target class or a non-target class, the input data set being biased in favor of the records of the non-target class, the decision tree comprising a plurality of nodes that include a root node and leaf nodes, said method comprising the steps of:

constructing the decision tree from the input data set, including the step of partitioning each of the plurality of nodes of the decision tree, beginning with the root node, based upon multivariate subspace splitting criteria;

computing distance functions for each of the leaf nodes;

identifying, with respect to the distance functions, a nearest neighbor set of nodes for each of the leaf nodes based upon a respective closeness of the nearest neighbor set of nodes to a target record of the target class; and

classifying and scoring the records, based upon the decision tree and the nearest neighbor set of nodes;

wherein said partitioning step comprises the step of:

for a current leaf node from among the leaf nodes of the decision tree,

computing a lowest value of a gini index achieved by univariate-based partitions on each of a plurality of attribute lists included in the current leaf node; and

wherein said partitioning step further comprises the steps of:

detecting subspace clusters of the records of the target class associated with the current leaf node;

computing the lowest value of the gini index achieved by distance-based partitions on each of the plurality of attribute lists included in the current leaf node, the distance-based partitions being based on distances to the detected subspace clusters;

partitioning pre-sorted attribute lists included in the current node into two sets of ordered attribute lists based upon a greater one of the lowest value of the gini index achieved by univariate partitions and the lowest value of the gini index achieved by distance-based partitions; and

creating new child nodes for each of the two sets of ordered attribute lists; and

wherein each of the plurality of pre-sorted attribute lists comprises a plurality of entries, and said step of partitioning the pre-sorted attribute lists comprises the steps of:

determining whether univariate partitioning or distance-based partitioning has occurred;

creating a first hash table that maps record ids of any of the records that satisfy a condition A=v to a left child node and that maps the record ids of any of the records that do not satisfy the condition A=v to a right child node, A being an attribute and v denoting a splitting position, when the univariate partitioning has occurred;

creating a second hash table that maps the record ids of any of the records that satisfy a condition Dist(d, p, w)=v to a left child node and that maps the record ids of any of the records that do not satisfy the condition Dist(d, p, w)=v to a right child node, when the distance-based partitioning has occurred, d being a record associated with a current subspace cluster, p being a centroid of the current subspace cluster, and w being a weight on dimensions of the current subspace cluster;

partitioning the pre-sorted attribute lists into the two sets of ordered attribute lists, based on information in a corresponding one of the first hash table or the second hash table;

appending each entry of the two sets of ordered attribute lists to one of the left child node or the right child node, based on the information in the corresponding one of the first hash table or the second hash table and information corresponding to the each entry, to maintain attribute ordering in the two sets of ordered attribute lists that corresponds that in the pre-sorted attribute lists.

34. (Previously Presented): A method for building a decision tree from an input data set, the input data set comprising records and associated attributes, the attributes including a class label attribute for indicating whether a given record is a member of a target class or a non-target class, the input data set being biased in favor of the records of the non-target class, the decision tree comprising a plurality of nodes that include a root node and leaf nodes, said method comprising the steps of:

constructing the decision tree from the input data set, including the step of partitioning each of the plurality of nodes of the decision tree, beginning with the root node, based upon multivariate subspace splitting criteria;

computing distance functions for each of the leaf nodes;

identifying, with respect to the distance functions, a nearest neighbor set of nodes for each of the leaf nodes based upon a respective closeness of the nearest neighbor set of nodes to a target record of the target class; and

classifying and scoring the records, based upon the decision tree and the nearest neighbor set of nodes;

wherein said classifying and scoring step comprises the steps of:

for each of the plurality of nodes of the decision tree, starting at the root node,

evaluating a Boolean condition and following at least one branch of the decision tree until a leaf node is reached;

classifying the reached leaf node based on a majority class of any of the predetermined attributes included therein;

for each node in the nearest neighbor set of nodes for the reached leaf node,

computing a distance between a record to be scored and a centroid of the reached leaf node, using a distance function computed for the reached leaf node; and

scoring the record using a maximum value of a score function, the score function defined as conf/dist(d,p,w,), wherein the conf is a confidence of the reached node, d is a particular record associated with a current subspace cluster, p is a centroid of the current subspace cluster, and w is a weight on dimensions of the subspace cluster.